arXiv:1508.06867v2 [stat.ME] 10 Dec 2015

Doubly robust goodness-of-fit test of coarse structural nested mean models with application to initiating combination antiretroviral treatment in HIV-positive patients

Shu Yang and Judith J. Lok

Abstract

Coarse structural nested mean models provide a useful tool to estimate treatment effects from longitudinal observational data with time-dependent confounders. However there is no existing guidance to specify the treatment effect model, and model misspecification can lead to biased estimators, preventing valid inference. To test whether the treatment effect model matches the data well, we derive a goodness-of-fit test procedure based on overidentification restrictions tests. We show that our test statistic is doubly-robust in the sense that with a correct treatment effect model, the test statistic has the correct level if either the treatment initiation model or a nuisance regression outcome model is correctly specified. In a simulation study we show the test procedure has correct type-I error and is powerful to detect model misspecification. In addition, we apply the test procedure to study how the timing of combination antiretroviral treatment initiation after infection predicts the one year treatment effect in HIV-positive patients with acute and early infection.

Keywords and phrases: Causal inference; Censoring; Doubly robust; Estimating Equation; Goodness-of-fit test; HIV/AIDs; Longitudinal observational study; Overiden-

tification restrictions test.

1 Introduction

The gold standard for evaluating effects of interventions are randomized controlled trials. However, they are not always available; for example, for evaluating a treatment strategy for HIV-positive patients, a randomized controlled trial would force patients to take the treatment or to be off the treatment regardless of their health status. Observational studies are useful in these settings. In observational studies, there often is time-dependent confounding by indication: some covariates are predictors of both the subsequent treatment and outcome, and are also affected by the past treatment history. Then, standard methods adjusting for the covariates history are fallible and can lead to bias (Robins et al., 1992; Robins, 2000; Robins et al. 2000).

Coarse structural nested mean models (Robins, 1998a) provide a useful tool to estimate treatment effects from longitudinal observational data. Lok and DeGruttola (2012) developed a time-dependent version of coarse structural nested mean models and applied it to investigate the impact of the timing of combination antiretroviral treatment initiation on the effect of one year treatment in HIV-positive patients. Their semiparametric method leads to an infinite number of unbiased estimating equations and a huge class of consistent and asymptotically normal estimators. An optimal estimator can be derived within this class of coarse structural nested mean models under well-specified models for the treatment effect, treatment initiation, and a nuisance regression outcome model, in an unpublished 2014 technical report available from the second author. The key assumption lies in a well-specified model for the treatment effect. However, no guidance exists on how to specify the treatment effect model, and model misspecification may lead to biased estimators, preventing valid inference.

The main contribution of this article is to derive a goodness-of-fit test statistic for testing

correct specification of the treatment effect model. The key insight is that with a correctly-specified treatment effect model we have more unbiased estimating equations than the number of parameters, which results in overidentification of the parameters. Overidentification restrictions tests, also called Sargan tests or J-tests (Sargan, 1958 and Hansen, 1982), are widely used in the econometric literature; they however seem to have been previously unnoticed in the biostatistics literature. The standard overidentification restrictions test, given by the minimized value of the generalized method of moments (Newey and McFadden, 1994; Imbens et al., 1995) criterion function, has a chi-squared limiting distribution, with degrees of freedom equal to the number of overidentification restrictions. In most situations, the minimum of the generalized method of moments criterion is obtained by a continuous iterative procedure to update the parameter estimates until convergence (Hansen et al., 1996). Arellano and Bond (1991) showed the test statistic based on one-step estimates other than the optimal generalized method of moments estimates is not robust and tends to over-reject even in large samples. Our test procedure is different from the standard overidentification restrictions tests in this regard. We do not obtain parameter estimates by minimizing an objective function, but rather we obtain parameter estimates by solving the optimal estimating equations with the number of equations equal to the number of parameters. The overly identified restrictions are only used for testing, not for estimation. This difference allows us to greatly reduce the computation burden. Our simulation studies show that our test statistic has correct size for large samples under the scenarios we considered. Another merit of the overidentification restrictions test is that no bootstrap is needed to compute the test statistic, which could be valuable with the large samples that are increasingly common.

2 Motivating problem and basic setup

2.1 The motivating problem

Combination antiretroviral treatment is the standard initial treatment for HIV, and has considerably reduced the morbidity and mortality in HIV-positive patients. In the HIV literature, findings imply that there are key early events, during acute and early infection, in the pathogenesis of HIV infection that determine the long-term pace of disease progression (Hecht et al., 2006). However, there is no strong evidence to support when to start treatment in patients in the acute and early stages of infection. It is important to understand the effect of initiating treatment at different times during the course of HIV infection. This investigation relies on an observational study, where we emulate a counterfactual experiment using causal models.

2.2 The Acute Infection and Early Disease Research Program

The Acute Infection and Early Disease Research Program study is a multicenter, observational cohort study of 1762 HIV-positive patients diagnosed during acute and early infection (Hecht et al., 2006). Dates of infection were estimated based on a stepwise algorithm that uses clinical and laboratory data (Smith et al., 2006). We included patients with CD4 and viral load measured within 12 months of the estimated date of infection, which resulted in 1696 patients. Let m denote the number of months between the estimated date of infection and combination antiretroviral treatment initiation ($m = 0, \ldots, 11$), where 0 indicates the estimated date of infection. We are interested in evaluating the impact of m on the effect of one year treatment.

2.3 Notation

Let Y_k be the patient's CD4 count at month k since the estimated date of infection ($k = 0, \ldots, K+1 \equiv 24$), and L_m be a vector of covariates measured at month m, including age,

gender, race, injection drug use, CD4 count and viral load. Let A_m be one if the patient was on treatment at month m and zero otherwise. We assume that once treatment is started, it is never discontinued. We use overbars to denote a variable history; for example, $\bar{A}_m$ is the treatment history until month m. Let T be the actual treatment initiation time. The patients are assumed to be an independent sample from a larger population (Rubin, 1978), and for notational simplicity we drop the subscript i for patients. To handle missingness, for L, if this is missing at month m, L_m was coded as "missing". For intermediate missing outcomes, we imputed Y_k by interpolation; if the outcome is missing just prior to onset of treatment, we imputed Y_k by carrying the last observation forward. Let $X \equiv (\bar{A}_K, \bar{L}_K, \bar{Y}_{K+1})$ denote the patient's full record. Until Section 6 we assume all patients are followed up until month $K+1$.

Let $Y_k^{(m)}$ be the CD4 count at month k, possibly counterfactual, had the patient started treatment at month m. Let $Y_k^{(\infty)}$ be the CD4 count at month k had the patient never started treatment during the course of follow up. We assume the patient's observed outcome Y_k is equal to the potential outcome $Y_k^{(m)}$ for m equal to the actual treatment initiation time T; that is, if $k > T$, $Y_k = Y_k^{(T)}$ and if $T \geq k$, $Y_k = Y_k^{(\infty)}$.

We assume the assumption of no unmeasured confounding (Robins et al., 1992):

$$Y_k^{(\infty)} \amalg A_m \mid \bar{L}_m, \bar{A}_{m-1} \ (k = m+1, \ldots, m+12), \tag{1}$$

where $\amalg$ denotes "is independent of" (Dawid, 1979). This assumption holds if $\bar{L}_m$ contains all prognostic factors for $Y_k^{(\infty)}$ that affect the treatment decision at month m. For example, if patients with lower CD4 counts initiated treatment earlier, the assumption (1) would fail to hold if $\bar{L}_m$ does not include the history of the CD4 count.

2.4 Coarse structural nested mean model

We model the treatment effect, comparing treatment starting at month m to never starting among the subgroup of patients with covariate history $\bar{l}_m$ and $T = m$, as

$$E\{Y_k^{(m)} - Y_k^{(\infty)} \mid \bar{L}_m^{(\infty)} = \bar{l}_m, T = m\} = \gamma_{m,\psi}^k(\bar{l}_m) \ (k = m, \ldots, m+12), \tag{2}$$

where ψ is the parameter in the treatment effect model. From now on, we consider $\gamma_{m,\psi}^k(\bar{l}_m) = (\psi_1 + \psi_2 m)(k-m)1_{(m \le k)}$, with $(k-m)$ the duration of treatment from month m to month k. We restrict the range of k from 12 to $K+1$, whereby we avoid making extra modeling assumptions beyond the necessary ones to estimate $\gamma_m^{m+12}(\bar{l}_m)$ in order to gain robustness. Particularly, $\gamma_{m,\psi}^{m+12}(\bar{l}_m)$ quantifies the effect of one year treatment if HAART was initiated at month m, among the subgroup of patients with covariate history $\bar{l}_m$. If outcome is the CD4 count and $\gamma_{m,\psi}^{m+12}(\bar{l}_m) > 0$, the effect of one year treatment is beneficial. $12\psi_1$ quantifies the effect of one year treatment if treatment was started at the estimated date of infection, and ψ_2 quantifies the increase of the effect of treatment for each month of delay after the estimated date of infection. Under this model, the treatment effect is homogeneous. In practice, the treatment effect may vary among different groups; for example, male and female patients may have different responses to the combination antiretroviral treatment. We can then extend the model as $\gamma_{m,\psi}^k(\bar{l}_m) = (\psi_1 + \psi_2 m + \psi_3 \text{gender})(k-m)1_{\{m \le k\}}$, where ψ_3 quantifies the magnitude and direction of the impact of gender.

For $k = 12, \ldots, K+1$, define $H(k) = Y_k - \gamma_T^k(\bar{L}_T)$. As proved in Robins et al. (1992), Lok et al. (2004) and Lok and DeGruttola (2012), $H(k)$ is mimicking a counterfactual outcome $Y_k^{(\infty)}$ in the sense that for $k = 12, \ldots, K+1$ and $m = k-12, \ldots, k-1$,

$$E\{H(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}, A_m\} = E\{Y_k^{(\infty)} \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}, A_m\}, \tag{3}$$

since by subtracting from the observed Y_k the average effect of treatment, we would obtain

the quantity that mimics the outcome had the patient not been treated. The implication of (3) and the assumption of no unmeasured confounding is that for $k = 12, \ldots, K+1$ and $m = k-12, \ldots, k-1$,

$$E\{H(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}, A_m\} = E\{H(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}, \tag{4}$$

which plays a key role for estimation.

2.5 The conditional probability of treatment initiation

We use a pooled logistic regression to model the probability of treatment initiation at month m, conditional on the past history, $p_\theta(m) \equiv P(A_m = 1 \mid \bar{A}_{m-1} = \bar{0}, \bar{L}_m; \theta) = 1_{(\bar{A}_{m-1}=\bar{0})} 1_{\text{visit}}(m)/[1 + \exp\{-\theta^T f(\bar{L}_m)\}]$, where $1_{\text{visit}}(m)$ is an indicator of whether a visit took place at month m, and $f(\bar{L}_m)$ is some function of $\bar{L}_m$. Let $J_{\text{trt}(\theta)}(X)$ denote the estimating function for θ_0.

2.6 The unbiased estimating equation and optimal estimation

Model (2) cannot be fit by standard regression methods because it involves potential outcomes. However, one can get consistent estimates by constructing unbiased estimating equations based on (4) (Lok and DeGruttola, 2012): for any measurable, bounded function $q_m^k : \bar{\mathcal{L}}_m \to \mathbb{R}^p$, $k = 12, \ldots, K+1$, $m = k-12, \ldots, k-1$, let

$$G_{(\psi,\theta,q)}(X) \equiv \sum_{k=12}^{K+1} \sum_{m=k-12}^{k-1} q_m^k(\bar{L}_m)[H_\psi(k) - E\{H_\psi(k)|\bar{L}_m, \bar{A}_{m-1} = \bar{0}\}]\{A_m - p_\theta(m)\}.$$

We use empirical process notation throughout. We let P denote the probability measure induced by X and let P_n denote the empirical measure induced by $X_1, ..., X_n$. Given a measurable function $f : \mathcal{X} \mapsto \mathbb{R}$, we write $P_n f(X) = n^{-1} \sum_{i=1}^n f(X_i)$ and $Pf(X)$ for the

expectation under P. Then

$$P_n\{\ G_{(\psi,\theta,q)}(X)^T \quad J_{\mathrm{trt}(\theta)}\ (X)^T\}^T = 0 \tag{5}$$

are the stacked unbiased estimating equations for both the parameter ψ and the (nuisance) parameter θ. For simplicity, we will suppress the dependence of the estimating functions on X; for example, $P_n G_{(\psi,\theta,q)}$ is shorthand for $P_n G_{(\psi,\theta,q)}(X)$. Sometimes, we also drop the dependence on the parameters.

In theory, q can be chosen arbitrarily; however it largely influences the precision of the resulting estimator. To derive the optimal estimating equation, and therefore the optimal estimator, we assume that for $k, s = 12, \ldots, K+1$ and m with $m = \max(k-12, s-12), \ldots, \min(k-1, s-1, K)$,

$$\mathrm{cov}\{H(k), H(s) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}, A_m\} \amalg A_m \mid \bar{L}_m, \bar{A}_{m-1}. \tag{6}$$

This assumption is a natural extension of (4). This would be true under Robins (1998b)'s rank preservation assumption and $(Y_k^{(\infty)}, Y_s^{(\infty)}) \amalg A_m \mid \bar{L}_m, \bar{A}_{m-1}$. However, the rank preservation assumption is unlikely to hold in practice and the assumption (6) is weaker.

The optimal estimating equations, within the class of $P_n G_{(\psi,\theta,q)}$ indexed by q for any measurable and bounded functions q_m^k, can be obtained by finding q^{opt} that satisfies $E\{\partial/\partial\psi^T G_{(\psi_0,\theta_0,q)}\} = E\{G_{(\psi_0,\theta_0,q)} G^T_{(\psi_0,\theta_0,q^{\mathrm{opt}})}\}$ for any q (Newey and McFadden, 1994). Then, under (6)

$$\begin{aligned} q_m^{\mathrm{opt}}(\bar{L}_m)^T &\equiv \{\mathrm{var}(H_m \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0})\}^{-1} \\ &\times \left\{ E\left(\frac{\partial}{\partial\psi} H_m \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}, A_m = 1\right) - E\left(\frac{\partial}{\partial\psi} H_m \mid \bar{L}_m, \bar{A}_m = \bar{0}\right)\right\}, \end{aligned} \tag{7}$$

where $q_m^{\mathrm{opt}} = (q_m^{\mathrm{opt},l}, \ldots, q_m^{\mathrm{opt},r})$ with $l = \max(m+1, 12)$ and $r = \min(m+12, K+1)$, which are informed by the fact that $m+1 \leq k \leq m+12$ and $12 \leq k \leq K+1$; $H_m = \{H_\psi(l), \ldots, H_\psi(r)\}^T$; $\mathrm{var}(H_m \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0})$ is a matrix with elements $\Gamma_{ks}^m \equiv$

$\text{cov}\{H_\psi(k), H_\psi(s) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$; and $E(\partial/\partial\psi H_m \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}, A_m = 1) - E(\partial/\partial\psi H_m \mid \bar{L}_m, \bar{A}_m = \bar{0}) = (\Delta_m^l, \ldots, \Delta_m^r)^T$ with $\Delta_m^k \equiv E\{\partial/\partial\psi H_\psi(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}, A_m\} - E\{\partial/\partial\psi H_\psi(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$.

Remark 1 *For the optimal estimating equations, defined by (5) and (7), we address two issues. One issue is that* $E\{H_\psi(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$ *and* q^{opt} *depend on* ψ. *Following Lok and DeGruttola (2012), we use a preliminary consistent estimate* $\hat{\psi}_p$ *to replace* ψ_0 *in* $E\{H_\psi(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$ *and* q^{opt}. *The rationale for this replacement is that the estimating equations are unbiased for any fixed value of* $\hat{\psi}_p$ *when* $p_\theta(m)$ *is correct. If* $\gamma_{m,\psi}^k$ *is linear in* ψ, Δ_m^k *does not depend on* ψ. *One choice of* $\hat{\psi}_p$ *is the optimal estimator if* q_m^k *is only non-zero for* $k = m + 12$. *Another issue is that* Δ_m^k *and* $E\{H_\psi(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$, *even with* ψ *in lieu of* ψ_p, *depend on the true unknown distribution. We will use parametric models to approximate these quantities. Let* $E\{H_{\psi_p}(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$ *be parametrized by* ξ_1; *for example* $E_{\xi_1}\{H_{\psi_p}(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$ *is a linear regression model with covariates* $\bar{L}_m$. *Likewise, let* Δ_m^k *be parametrized by* ξ_2. *Denote estimating functions for* ξ_1, ξ_2 *and* ψ_p *by* $J_{1(\xi_1,\psi_p)}$, $J_{2(\xi_2)}$ *and* $G_{p(\psi_p,\xi_2)}$. *Since* G_p *depends on* Δ_m^{m+12}, *it is also a function of* ξ_2.

3 Asymptotic results of optimal estimators

We present the consistency and asymptotic normality result of the optimal estimator. These results are the building blocks to derive the goodness-of-fit test statistic.

Theorem 1 (Consistency) *Let* $G^*_{(\psi,\psi_p,\xi,\theta)}$ *be optimal estimating functions*

$$G^*_{(\psi,\psi_p,\xi,\theta)} = \sum_{k=12}^{K+1} \sum_{m=k-12}^{k-1} q^{k,\text{opt}}_{m,\psi_p,\xi_2}(\bar{L}_m)[H_\psi(k) - E_{\xi_1}\{H_{\psi_p}(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}]\{A_m - p_\theta(m)\},$$

and $U_{(\psi,\psi_p,\xi,\theta)} = \{ \; G^*_{(\psi,\psi_p,\xi_1,\xi_2,\theta)} \quad G_{p(\psi_p,\xi_2)} \quad J_{1(\xi_1,\psi_p)} \quad J_{2(\xi_2)} \quad J_{\text{trt}(\theta)}\}^T$ *be a system of estimating functions stacking all estimating functions together, where* G_p, J_1 *and* J_2 *are defined*

in Remark 1, and J_{trt} *is defined in (5). Let* $(\hat{\psi}, \hat{\psi}_p, \hat{\xi}, \hat{\theta})$ *be the solution to estimating equations* $P_n U_{(\psi,\psi_p,\xi,\theta)} = 0$*. The true parameter values are* ψ_0*,* ξ_0 *and* θ_0*. Under the regularity conditions (C1)–(C2) specified in the Supplementary Material, if the treatment effect model* $\gamma^k_{m,\psi}(\bar{L}_m)$ *is well specified, and either* $E_{\xi_1}\{H_\psi(k)|\bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$ *or* $p_\theta(m)$ *is well specified,* $\hat{\psi} - \psi_0 \to 0$ *in probability, as* $n \to \infty$*.*

Theorem 2 (Asymptotic normality) *Under the regularity conditions (C1)–(C5) specified in the Supplementary Material,* $n^{1/2}(\hat{\psi} - \psi_0) \to N_p(0, \Sigma_\psi)$ *in distribution, as* $n \to \infty$*, where* p *is the dimension of* ψ_0 *and* Σ_ψ *is the* $p \times p$ *upper left matrix in* $\{P\partial/\partial(\psi, \psi_p, \xi, \theta)U\}^{-1}P(U U^T)\{P\partial/\partial(\psi, \psi_p, \xi, \theta)U\}^{-1T}$*.*

Remark 2 *In the statistics literature, estimators solving unbiased estimating equations are often called Z-estimators. The theory of consistency and asymptotic normality of Z-estimators is well established, see for example Theorem 5.9 and Section 5.3 in Van der Vaart (2000). We skip the detailed proof but explain the regularity conditions needed to guarantee the consistency in the Supplementary Material. From Theorem 1, the functional form of* $\gamma^k_{m,\psi}$ *must be correctly specified. In contrast, the estimator remains consistent for* ψ *if either* $E_{\xi_1}\{H_\psi(k)|\bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$ *or* $p_\theta(m)$ *is well specified, but not necessary both. The estimator is doubly robust (Robins and Rotnitzky, 2001; Van der Laan and Robins, 2003). The functional form of the nuisance models can be selected on the basis of the observed data, as well as the literature and subject knowledge specific to the application setting. Later in this article, we provide a more specific illustration in the context of our example.*

4 Goodness-of-fit test

The consistency, asymptotic normality, and double robustness of estimators rely on a key assumption, that is, the treatment effect model is well specified. Misspecification of the treatment effect model causes bias in the parameter estimation and break down the asymptotic results. We now develop tests for model specification based on overidentification restrictions

tests. Conceptually, for a well specified model, a new set of unbiased estimating functions, other than the optimal ones that are used for estimation, evaluated at the optimal estimators, should be asymptotically concentrated at zero. This asymptotic behavior leads to the following theorem:

Theorem 3 (Goodness-of-Fit Test) *Let the treatment effect model be* $\gamma^k_{m,\psi}(\bar{l}_m)$ *and* $H_\psi(k) = Y_k - \gamma^k_{T,\psi}(\bar{l}_T)$. *Choose a set of functions* $\{\tilde{q}^k_m(\bar{L}_m) \in \mathbb{R}^\nu, k = 12, \ldots, K+1, m = k-12, \ldots, k-1\}$ *that are different from the optimal choice* $q^{k,\text{opt}}_m$. *Let*

$$\tilde{G}_{(\psi,\psi_p,\xi,\theta)} = \sum_{k=12}^{K+1} \sum_{m=k-12}^{k-1} \tilde{q}^k_{m,\xi_2}(\bar{L}_m)[H_\psi(k) - E_{\xi_1}\{H_{\psi_p}(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}]\{A_m - p_\theta(m)\}.$$

Let $(\hat{\psi}, \hat{\psi}_p, \hat{\xi}, \hat{\theta})$ *be as in Theorem 1. The null hypothesis is* H_0*:* $\gamma^k_m(\bar{l}_m)$ *is well specified, and either* $E_{\xi_1}\{H_\psi(k) \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$ *or* $p_\theta(m)$ *is well specified. Under* H_0 *and the regularity conditions (C1)–(C10) specified in the Supplementary Material, the goodness-of-fit test statistic*

$$GOF = n\{P_n\tilde{G}_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})}\}^T\hat{\Sigma}^{-1}P_n\tilde{G}_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})} \to \chi^2(\nu), \tag{8}$$

in distribution, as $n \to \infty$, *where* Σ *is the variance of* $\Phi_{(\psi_0,\psi_0,\xi_0,\theta_0)}$ *with* $\Phi_{(\psi_0,\psi_0,\xi_0,\theta_0)}$ *the asymptotic linear representation of* $\tilde{G}_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})}$, *which is a linear combination of* G^*, $\tilde{G}$, G_p, J_1, J_2 *and* J_{trt}, *defined in (6) in the Supplementary Material, and* $\hat{\Sigma}$ *is the sample variance of* $\Phi_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})}$.

We state the key steps in the proof, with details in the Supplementary Material. To establish the asymptotic distribution of $n^{1/2}P_n\tilde{G}_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})}$, and therefore that of GOF, a key step is to linearise $n^{1/2}P_n\tilde{G}_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})}$ as $n^{1/2}P_n\Phi_{(\psi_0,\psi_0,\xi_0,\theta_0)}$ for some Φ, whereby we can apply the typical central limit theorem. To do so, the Lipschitz condition (C7) implies the functions $\tilde{G}_{(\psi,\psi_p,\xi,\theta)}$ form a Donsker class. Using Lemma 19.24 of Van der Vaart (2000), we have

$$\sqrt{n}(P_n - P)\tilde{G}_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})} \;=\; \sqrt{n}(P_n - P)\tilde{G}_{(\psi_0,\psi_0,\xi_0,\theta_0)} + o_p(1). \tag{9}$$

Next, we apply a Taylor expansion to $P\tilde{G}_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})}$ on the left hand side of (9) and use the fact that $P\tilde{G}_{(\psi_0,\psi_0,\xi_0,\theta_0)} = 0$ on the right hand side of (9). Finally, we can express $\Phi_{(\psi_0,\psi_0,\xi_0,\theta_0)}$, the asymptotic linear representation of $\tilde{G}_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta})}$, to be a linear combination of G^*, $\tilde{G}$, G_p, J_1, J_2 and J_{trt}.

Remark 3 (Double Robustness) *The goodness-of-fit test statistic is doubly robust in the sense that for (8) to hold we only require that either* $E_{\xi_1}\{H_\psi(k)|\bar{L}_m, \bar{A}_{m-1} = \bar{0}\}$ *or* $p_\theta(m)$ *is well specified, not necessary for both. This property adds a protection from possible misspecification of the nuisance models.*

Remark 4 *The standard overidentification restrictions test is* $\min_\psi nP_nV_{(\psi)}^T\{\hat{\Sigma}(\psi)\}^{-1}P_nV_{(\psi)}$, *where* $V_{(\psi)} \equiv (\ G^*_{(\psi,\hat{\psi}_p,\hat{\xi}_1,\hat{\xi}_2,\hat{\theta})}\ \ \tilde{G}_{(\psi,\hat{\psi}_p,\hat{\xi},\hat{\theta})}\ )^T$ *and* $\Sigma(\psi)$ *is the asymptotic variance of* $P_nV_{(\psi)}$. *In most situations, the minimum is obtained by a continuous iterative procedure to update the parameter estimates; that is,* $\hat{\psi}^{(t+1)} = \arg\min_\psi nP_nV_{(\psi)}^T\{\hat{\Sigma}(\hat{\psi}^{(t)})\}^{-1}P_nV_{(\psi)}$ *until convergence (Hansen et al., 1996). Our test procedure does not need any iterative procedure, which simplifies the calculation.*

Remark 5 (Choosing $\tilde{q}$) *Just like a naive choice of q in estimating equations may lead to an estimator with large variance and thus useless inference, an arbitrary choice of* $\tilde{q}$ *may lead to the goodness-of-fit test lacking of power. We propose the following procedure to choose* $\tilde{q}$, *which is powerful in certain circumstances. Suppose we have two models to choose from for the treatment effect. Let the null model be* γ^*_ψ, *which is the treatment effect model we are testing for, and the other model to be an alternative model* $\tilde{\gamma}_\psi$. *We can derive* Δ^*, $q^{*\text{opt}}$, $\tilde{\Delta}$, *and* $\tilde{q}^{\text{opt}}$ *as in (7) with* γ^*_ψ *and* $\tilde{\gamma}_\psi$. *Note that* $q^{*\text{opt}}$ *is used for optimal estimation of the parameters in the null model. Then, candidates for* $\tilde{q}$ *are* Δ^*, $\tilde{\Delta}$, $\tilde{q}^{\text{opt}}$, *or any subvector of these that is not included in* $q^{*\text{opt}}$. *Our simulation study shows that the goodness-of-fit test with* $\tilde{q}^{\text{opt}}$ *is most powerful among this set of candidates in detecting the alternative model.*

5 Extension of goodness-of-fit test in the presence of censoring

We use the Inverse-Probability-of-Censoring-Weighting technique (Robins et al., 1995; Hernán et al., 2005; Lok and DeGruttola, 2012) to accommodate patients lost to follow-up. Let $C_p = 0$ indicate a patient remains in the study at month p. Following Lok and DeGruttola (2012), we assume that censoring is missing at random; that is, $(\bar{L}, \bar{A}) \amalg C_{k+1} \mid \bar{L}_k, \bar{A}_k, \bar{C}_k = \bar{0}$, whereby we have $P(A_m = 1 \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0}, \bar{C}_m = \bar{0}) = P(A_m = 1 \mid \bar{L}_m, \bar{A}_{m-1} = \bar{0})$, and $p_\theta(m)$ does not depend on censoring. Define the Inverse-Probability-of-Censoring-Weightingg version of estimating functions G^{*c} and $\tilde{G}^c$ using weights $W^k_{m,\eta} = 1/\{\prod_{p=m+1}^{k} P_\eta(C_p = 0 \mid \bar{L}_{p-1}, \bar{A}_{p-1}, \bar{C}_{p-1} = \bar{0})\}$, see the Supplementary Material for details. In calculation of the weights, we use a pooled logistic regression model to estimate $P_\eta(C_p = 0 \mid \bar{L}_{p-1}, \bar{A}_{p-1}, \bar{C}_{p-1} = \bar{0})$. We assume the censoring model is well specified with estimating functions $J_{\text{cen}(\eta)}$. Similarly, we have the Inverse-Probability-of-Censoring-Weighting version of the estimating function for the preliminary estimator $\hat{\psi}_p$, denoted by G^c_p. For the nuisance regression outcome models, the regression was restricted to patients still in follow-up and use weighted regression analysis with the censoring weights.

Define the goodness-of-fit test statistic as

$$GOF^c = n\{P_n \tilde{G}^c_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta},\hat{\eta})}\}^T (\hat{\Sigma}^c)^{-1} P_n \tilde{G}^c_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta},\hat{\eta})},$$

where $\hat{\Sigma}^c$ is the sample variance of $\Phi^c_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta},\hat{\eta})}$, with $\Phi^c_{(\psi_0,\psi_0,\xi_0,\theta_0,\eta_0)}$ the asymptotic linear representation of $\tilde{G}^c_{(\hat{\psi},\hat{\psi}_p,\hat{\xi},\hat{\theta},\hat{\eta})}$, defined by (10) in the Supplementary Material. As proved in the Supplementary Material, subject to regularity conditions, GOF^c has an asymptotic chi-squared distribution, with degrees of freedom dimension of $\tilde{G}^c$.

6 Simulations

The simulation designs were based on the Acute Infection and Early Disease Research Program database, but we did not consider censoring. Following an unpublished 2014 technical report available from the second author, we first generated the CD4 count outcomes $Y_k^{(\infty)}$ under no treatment, followed by a treatment initiation time T, and lastly the observed outcomes Y_k ($k = 6, \ldots, 30$), as follows: (i) In each sample, 2 groups were simulated: injection drug users (10%) and patients who never used drugs (90%), and then $\log Y_6^{(\infty)} \sim N(6 \cdot 0, 0 \cdot 4^2)$ for injection drug users, and $N(6 \cdot 6, 0 \cdot 5^2)$ for non injection drug users. For $k \geq 6$, $Y_{k+1}^{(\infty)} = -10 + Y_k^{(\infty)} + \epsilon_{k+1}$, where $\epsilon_k \sim N(0, \sigma_k^2)$ with $\sigma_k = 52 \cdot 375 - 1 \cdot 625k$ for $k = 7, \ldots, 19$ and $\sigma_k = 21 \cdot 5$ for $k = 20, \ldots, 30$; (ii) T was generated by a logistic regression model $\text{logit}\{P(T = m \mid T \geq m, \bar{L}_m)\} = -2 \cdot 4 - 0 \cdot 42\text{injdrug} - 0 \cdot 0035 Y_m^{(\infty)} - 0 \cdot 026m$, where injdrug is an indication of being an injection drug user; and (iii) $Y_k = Y_k^{(\infty)} + \gamma_T^k(\bar{L}_T)$. We considered different models for γ_m^k.

The performance of the test statistics was assessed by their ability (i) to confirm the adequacy of a model that is correctly specified with the data-generating model (type-I error) and (ii) to reject a misspecified model (power). The model under the null hypothesis H_0 upon which the goodness-of-fit statistic is based, and the alternative hypothesis H_a were specified as $H_0 : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m)(k - m)$ versus $H_a : \gamma_{m,\psi}^k \neq (\psi_1 + \psi_2 m)(k - m)$. Six scenarios regarding the true treatment effect model, H_0 and a parametric specification of H_a were specified as follows:

Scenario (a): True: $\gamma_{m,\psi}^k = (25 - 0 \cdot 7m)(k - m)$, $H_0 : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m)(k - m)$, and $H_a : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m + \psi_3 m^2)(k - m)$;

Scenario (b): True: $\gamma_{m,\psi}^k = (25 - 0 \cdot 7m)(k - m)$, $H_0 : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m + \psi_3 \text{injdrug})(k - m)$, and $H_a : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m + \psi_3 m^2)(k - m)$;

Scenario (c): True: $\gamma_{m,\psi}^k = (35 - 1 \cdot 1m + 0 \cdot 04m^2)(k - m)$, $H_0 : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m)(k - m)$, and $H_a : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m + \psi_3 m^2)(k - m)$;

Scenario (d): True: $\gamma_{m,\psi}^k = (35 - 1 \cdot 1m + 0 \cdot 04k^2)(k - m)$, $H_0 : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m)(k - m)$,

and $H_a : \gamma^k_{m,\psi} = (\psi_1 + \psi_2 m + \psi_3 m^2)(k-m)$;

Scenario (e): True: $\gamma^k_{m,\psi} = (25 - m + 0 \cdot 03m^2)(k-m)$, $H_0 : \gamma^k_{m,\psi} = (\psi_1 + \psi_2 m)(k-m)$, and $H_a : \gamma^k_{m,\psi} = (\psi_3 + \psi_4 m)(k-m)^{3/2}$;

Scenario (f): True: $\gamma^k_{m,\psi} = (10 - 1 \cdot 1m)(k-m)^{3/2}$, $H_0 : \gamma^k_{m,\psi} = (\psi_1 + \psi_2 m)(k-m)$, and $H_a : \gamma^k_{m,\psi} = (\psi_3 + \psi_4 m)(k-m)^{3/2}$.

Specifically, in Scenarios (a) and (b), H_0 is correctly specified. In Scenarios (c)–(f), H_0 is misspecified with different degrees of departure from the true model. In Scenarios (c) and (d), H_0 is nested in the parametric specification of H_a. In Scenarios (e) and (f), H_0 is not nested in the parametric specification of H_a.

Type-I error and power were estimated by the frequency of rejecting H_0 using $1,000$ simulated datasets. We considered the following choices of $\tilde{q}$: (i) $\tilde{q}^k_m \equiv 1$, which is a naive choice for comparison; (ii) $\tilde{q}^k_m = \tilde{\Delta}^k_m$; and (iii) $\tilde{q}^k_m = \tilde{q}^{\text{opt},k}_m$. (ii) and (iii) were derived under the parametric specification of H_a.

The optimal estimator was obtained by solving (5) with (7). In (5), the treatment initiation model was fitted by a logistic regression model adjusting for Y_m, injection drug use, and month, restricted to patients and visits with $\bar{A}_{m-1} = \bar{0}$. Thus, the treatment initiation model was correctly specified. $E\{H_{\psi_p}(k) \mid \bar{A}_{m-1} = \bar{0}, \bar{L}_m\}$ was fitted by a linear regression model adjusting for CD4_m and $(k-m)$, restricted to patients and visits with $\bar{A}_{m-1} = \bar{0}$. The covariates were motivated by (3) and the data generating mechanism. The nuisance models in q^{opt}_m are specified in the Supplementary Material for simplicity of presentation, which do not affect the double robustness of the estimator.

In addition to the goodness-of-fit test statistic, we considered an elaborated-model-fitting-and-testing approach, which combines the null model and the parametric specification of the alternative model and tests the significance of the parameters corresponding to the alternative model.

From Scenarios (a) and (b) in Table 1, where the treatment effect model under H_0 is correctly specified, the goodness-of-fit test procedure with all choices of $\tilde{q}$ controls type-

I error for $n = 1,000$ and $n = 2,000$. This suggests that the chi-squared distribution derived in this article provides an accurate approximation to the finite sample behavior of the goodness-of-fit test statistic for these sample sizes.

From Scenarios (c)–(f), where the treatment effect model is not correctly specified, the goodness-of-fit test procedure with the optimal $\tilde{q}_m^k$ derived under the parametric specification of H_a is most powerful, and as the sample size increases, the power increases, confirming the theoretical results. From Scenarios (c) and (d), the goodness-of-fit test procedure and the elaborated-model-fitting-and-testing approach are comparable when testing nested models. In both scenarios, the goodness-of-fit test procedure is slightly more powerful than the elaborated-model-fitting-and-testing approach for $n = 500$ and $n = 1,000$, which is not apparent for $n = 2,000$. For Scenarios (e) and (f), the null treatment effect model is not nested in the parametric specification of H_a. Under Scenario (e), the goodness-of-fit test statistic with $\tilde{q}_m^{\text{opt},k}$ shows more power than the elaborated-model-fitting-and-testing approach, likely because the elaborated-model-fitting-and-testing approach fits a larger model and loses power. Under Scenario (f), the goodness-of-fit test statistic with $\tilde{q}_{3m}^{\text{opt},k}$ is slightly more powerful than the elaborated-model-fitting-and-testing approach for $n = 500$, and both approaches are powerful to reject the null model in other cases.

Table 1: Type-I error estimates and power estimates ($\times 100$) for testing the null model H_0 by the proposed goodness-of-fit (GOF) test statistic with $\tilde{q}$ being 1, $\tilde{\Delta}$, and $\tilde{q}^{\text{opt}}$, and the Elaborated Model Fitting and Testing (EMFT) approach over $1,000$ simulations under Scenarios (a)–(f)

Type-I error estimates in Scenario (a)

	GOF			EMFT
$n\backslash\tilde{q}$	1	$\tilde{\Delta}_m^k$	$\tilde{q}_m^{\text{opt},k}$	
500	$5\cdot3$	$4\cdot3$	$5\cdot2$	$4\cdot9$
1000	$4\cdot5$	$5\cdot7$	$5\cdot6$	$5\cdot4$
2000	$4\cdot8$	$4\cdot4$	$5\cdot2$	$5\cdot3$

Type-I error estimates in Scenario (b)

	GOF			EMFT
	1	$\tilde{\Delta}_{3m}^k$	$\tilde{q}_m^{\text{opt},k}$	
	$9\cdot1$	$9\cdot8$	$12\cdot3$	$13\cdot5$
	$5\cdot3$	$4\cdot4$	$5\cdot4$	$5\cdot7$
	$5\cdot2$	$4\cdot4$	$5\cdot2$	$5\cdot3$

Power estimates in Scenario (c)

	GOF			EMFT
$n\backslash\tilde{q}$	1	$\tilde{\Delta}_m^k$	$\tilde{q}_m^{\text{opt},k}$	
500	15	29	59	56
1000	28	55	89	84
2000	52	88	99	99

Power estimates in Scenario (d)

	GOF			EMFT
	1	$\tilde{\Delta}_m^k$	$\tilde{q}_m^{\text{opt},k}$	
	90	96	97	83
	100	100	100	97
	100	100	100	100

Power estimates in Scenario (e)

	GOF			EMFT
$n\backslash\tilde{q}$	1	$\tilde{\Delta}_m^k$	$\tilde{q}_m^{\text{opt},k}$	
500	12	25	49	28
1000	24	53	73	54
2000	48	79	91	80

Power estimates in Scenario (f)

	GOF			EMFT
	1	$\tilde{\Delta}_m^k$	$\tilde{q}_m^{\text{opt},k}$	
	93	99	100	96
	100	100	100	100
	100	100	100	100

7 Application

We applied the proposed goodness-of-fit test to study how the timing of combination antiretroviral treatment initiation after infection predicts the effect of one year of treatment in HIV-positive patients. We used the Acute Infection Early Disease Research Program database. We started with a simple null model for the treatment effect, $H_0 : \gamma_{m,\psi}^k = (\psi_1 + \psi_2 m)(k-m)$, and conducted directed alternative-model tests by testing whether possible effect modifiers should be added into the model. In the HIV literature, it has been found that there may be gender differences in immunologic response to combination antiretroviral treatment: early studies suggested that clinical disease progression was more rapid in women than men with combination antiretroviral treatment (Friedland et al., 1991; Bozzette et al., 1998); conversely, more recent studies have shown that women have better immunologic outcomes

than men on treatment(Maman et al., 2012; Maskew et al., 2013). It has also been shown that older age is associated with a poorer CD4 count increase with combination antiretroviral treatment (Maman et al., 2012; Maskew et al., 2013). Injection drug use has been found to be associated with reduced effectiveness of combination antiretroviral treatment (Poundstone et al., 2001). As suggested by the literature, we considered tests directed at 3 variables: gender, age, and injection drug use. For the test directed at a certain variable Z, we calculated the goodness-of-fit test statistic with $\tilde{q}$ being the optimal form derived from the parametric specification of the alternative model $\tilde{\gamma}^k_{m,\psi} = (\psi_1 + \psi_2 m + \psi_3 Z)(k-m)1_{(k>m)}$.

The nuisance models were specified on the basis of the observed data, the clinical literature, and subject knowledge. For the censoring model, we used a logistic regression model adjusting for square root of current CD4 count ($\text{CD4}_m^{1/2}$), current log viral load, gender, age, injection drug use (injdrug), month, squared month, and whether a patient was treated, as discussed in Krishnan et al. (2011) and Lok et al. (2010). For the treatment initiation model, we used a logistic regression models including $\text{CD4}_m^{1/2}$, current log viral load, gender, age, injection drug use, month, days since last visit, indication of first visit, indication of second visit, race, as discussed in Lok and Griner (2014). For $E_{\xi_2}\{H(k) \mid \bar{L}_m, \bar{A}_m = \bar{0}\}$, we used a regression model adjusting for CD4_m, $\text{CD4}_m^{3/4}(k-m)$, $\text{CD4}_m^{3/4}\text{age}(k-m)$, $\text{CD4}_m^{3/4}\text{race}(k-m)$, $\text{CD4}_m^{3/4}\text{injdrug}(k-m)$, whether there is a CD4 slope measure, $\text{CD4slope}_m(k-m)^{1/2}$, $(6-m)^+$, and $(6^2-m^2)^+$ with $a^+ \equiv a \times 1(a>0)$. The model was motivated by (3). The inclusion of CD4_m, $\text{CD4}_m^{3/4}(k-m)$, and $\text{CD4}_m^{3/4}\text{age}(k-m)$ was suggested from a stochastic model of $Y_{ik}^{(\infty)}$ for each patient i over time k, $\{Y_{ik}^{(\infty)}\}^{1/4} = a_i + bk + \gamma_1\text{age} + \gamma_2\text{age}k + \phi W_{ik} + \epsilon_{ik}$, where a_i is a normal random effect, W_{ik} is a Brownian motion process, ϵ_{ik} is normal with mean zero and constant variance, and a_i, W_{ik}, ϵ_{ik} and age are independent (Taylor et al., 1994). Other covariates were suggested in Taylor and Law (1998) and May et al. (2009).

Table 2 shows the results from fitting the optimal estimator of the null treatment effect model, along with the goodness-of-fit tests directed at gender, age, and injection drug use.

The p-values are all greater than $0 \cdot 05$. To avoid the multiple testing problem, we did not consider other tests. The three tests were specified prior to the actual calculation. The results show a benefit of combination antiretroviral treatment; for example, starting treatment at the estimated date of infection would lead to an expected added improvement in CD4 counts of $12\hat{\psi}_1 = 299$ cells/mm^3 after a year of therapy. Delaying treatment initiation during acute and early infection may diminish the CD4 count gain associated with one year treatment ($\hat{\psi}_2 < 0$); however, this result is not statistically significant.

Table 2: The Acute Infection Early Disease Research Program data: the optimal estimator fitting the null treatment effect model: point estimate (95% confidence intervals based on the asymptotic normality result), along with goodness-of-fit statistics (Statistic), associated degree of freedom (DF), and p-values (p-value) for the adequacy of the null model by testing whether gender or injection drug use should be added into the model

$\hat{\psi}_1$(95% CI)	$\hat{\psi}_2$ (95% CI)
$24 \cdot 88(21 \cdot 61, 28 \cdot 15)$	$-0 \cdot 48(-1 \cdot 47, 0 \cdot 52)$

Goodness-of-fit test			
	Statistic	DF	p-value
Test directed at gender	$0 \cdot 99$	1	$0 \cdot 32$
Test directed at age	$0 \cdot 80$	1	$0 \cdot 37$
Test directed at injection drug use	$2 \cdot 93$	1	$0 \cdot 09$

8 Discussion

The applicability of the goodness-of-fit test procedure presented in this article is broad in the causal inference literature. The testing procedure can also be developed for the traditional structural nested mean models (Robins, 1994) other than the time-dependent coarse structural nested mean models considered in this article, and marginal structural models (Robins, 2000), because both approaches yield overidentification of the parameters. For the Inverse-Probability-of-Censoring-Weighting estimator of marginal structural models, unbiased estimating equations are $P_n q(V)\{Y - \mu(\bar{A}, V)\}w(\bar{A} \mid \bar{L}) = 0$, where Y is the outcome at the end of study, $\bar{A}$ is the treatment history, V is a subset of the baseline covariates, $\mu(\bar{a}, V) \equiv E(Y^{\bar{a}} \mid V)$ is the marginal structural model, where $Y^{\bar{a}}$ is the counterfactual out-

come had every individual received the treatment $\bar{a}$, and $w(\bar{A} \mid \bar{L})$ is the inverse of conditional probability of receiving the actual treatment $\bar{A}$ given $\bar{L}$. These equations are unbiased for most choices of $q(V)$, leading to a large class of unbiased estimating equations. The literature of structural nested mean models and marginal structural models concentrates on estimation and efficiency. Little attention has been given to goodness-of-fit tests. Our test procedure for the treatment effect model can be developed in these contexts in the same manner.

Our goodness-of-fit test procedure can also deal with treatment of the form "initiate treatment when the CD4 count first drops below x". Especially in resource limited countries, and historically also in the US, initiation of combination antiretroviral treatment is decided based on the CD4 count threshold. Orellana et al. (2010) proposed dynamic regime marginal structural models and Lok et al. (2007) used structural nested mean models to simultaneously compare dynamic treatment regimes of this form and estimate the optimal one. Due to the popularity of these methods, the development of goodness-of-fit tests in these settings will be useful for model diagnosis and protect causal estimates from biases introduced by misspecification of the treatment effect model.

Acknowledgements

We are grateful to the patients who volunteered for AIEDRP, to the AIEDRP study team, and to Susan Little, Davey Smith, and Christy Anderson for their help and advice in interpreting the AIEDRP database. We would like to thank James Robins and Victor DeGruttola for insightful and fruitful discussions. This work was sponsored by the Milton Fund, the Career Incubator Fund award from the Harvard School of Public Health, and NIH grants NIAID R01AI100762, R01 AI51164, R37 AI032475, AI43638, AI74621, and AI36214.

Supplementary material

Supplementary material available at *Biometrika* online includes regularity conditions, the proof of Theorem 3, the derivation of the goodness-of-fit test statistic in the presence of censoring, and the nuisance regression outcome models used in the simulation.

References

Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Rev. econ. stud.*, 58(2):277–297, 1991.

Samuel A Bozzette, Sandra H Berry, Naihua Duan, Martin R Frankel, Arleen A Leibowitz, Doris Lefkowitz, Carol-Ann Emmons, J Walton Senterfitt, Marc L Berk, Sally C Morton, et al. The care of HIV-infected adults in the united states. *New Engl. J. Med.*, 339(26): 1897–1904, 1998.

A Philip Dawid. Conditional independence in statistical theory. *J. R. Statist. Soc. B*, 41: 1–31, 1979.

Gerald H Friedland, Brian Saltzman, Joan Vileno, Katherine Freeman, Lewis K Schrager, and Robert S Klein. Survival differences in patients with AIDS. *JAIDS J. Acq. Im. Deficien. Synd.*, 4(2):144–153, 1991.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: J. Economet. Soc.*, 50:1029–1054, 1982.

Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-sample properties of some alternative gmm estimators. *J. Busi. & Econ. Stat.*, 14(3):262–280, 1996.

Frederick M Hecht, Lei Wang, Ann Collier, Susan Little, Martin Markowitz, Joseph Margolick, J Michael Kilby, Eric Daar, and Brian Conway. A multicenter observational study

of the potential benefits of initiating combination antiretroviral therapy during acute HIV infection. *J. Infect. Dis.*, 194(6):725–733, 2006.

Miguel A Hernán, Stephen R Cole, Joseph Margolick, Mardge Cohen, and James M Robins. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharm. & Drug Saf.*, 14(7):477–491, 2005.

Guido Imbens, Phillip Johnson, and Richard H Spady. *Information theoretic approaches to inference in moment condition models*. National Bureau of Economic Research Cambridge, Mass., USA, 1995.

S Krishnan, K Wu, M Smurzynski, RJ Bosch, CA Benson, AC Collier, MK Klebert, J Feinberg, and SL Koletar. Incidence rate of and factors associated with loss to follow-up in a longitudinal cohort of antiretroviral-treated HIV-infected persons: an AIDS Clinical Trials Group (ACTG) Longitudinal Linked Randomized Trials (ALLRT) analysis. *HIV Clin. Tria.*, 12(4):190–200, 2011.

J. J. Lok, M. A. Hérnan, and J. M. Robins. Optimal start of HAART treatment in HIV positive patients. *Proceedings of the Joint Statistical Meetings*, pages 1149–1160, 2007.

Judith Lok, Richard Gill, Aad Van Der Vaart, and James Robins. Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statist. Neerland.*, 58(3):271–295, 2004.

Judith J Lok and Victor DeGruttola. Impact of time to start treatment following infection with application to initiating haart in HIV-positive patients. *Biometrics*, 68(3):745–754, 2012.

Judith J Lok and Ray Griner. Optimal estimation of coarse structural nested mean models with application to initiating haart in HIV-positive patients. *Submitted*, 2014.

Judith J Lok, Ronald J Bosch, Constance A Benson, Ann C Collier, Gregory K Robbins, Robert W Shafer, and Michael D Hughes. Long-term increase in CD4+ T-cell counts during combination antiretroviral therapy for HIV-1 infection. *AIDS (London, England)*, 24(12):1867–1876, 2010.

David Maman, Mar Pujades-Rodriguez, Fabien Subtil, Loretxu Pinoges, Megan McGuire, Rene Ecochard, and Jean-François Etard. Gender differences in immune reconstitution: a multicentric cohort analysis in sub-Saharan Africa. *PLoS One*, 7(2):e31078, 2012.

Mhairi Maskew, Alana T Brennan, Daniel Westreich, Lynne McNamara, A Patrick MacPhail, and Matthew P Fox. Gender differences in mortality and CD4 count response among virally suppressed HIV-positive patients. *J. Wom. Health.*, 22(2):113–120, 2013.

Margaret May, Robin Wood, Landon Myer, Patrick Taffé, Andri Rauch, Manuel Battegay, and Matthias Egger. CD4 T-cell declines by ethnicity in untreated HIV-1 infected patients in South Africa and Switzerland. *J. Infect. Dis.*, 200(11):1729–1735, 2009.

Whitney K Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook economet.*, 4:2111–2245, 1994.

Liliana Orellana, Andrea Rotnitzky, and James M Robins. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. *Internat. J. Biostat.*, 6(2):1557–79, 2010.

Katharine E Poundstone, Richard E Chaisson, and Richard D Moore. Differences in HIV disease progression by injection drug use and by sex in the era of highly active antiretroviral therapy. *AIDS*, 15(9):1115–1123, 2001.

James M Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Comm. Stat. Theo. Meth.*, 23(8):2379–2412, 1994.

James M Robins. Correction for non-compliance in equivalence trials. *Stat. in Med.*, 17(3): 269–302, 1998a.

James M Robins. Structural nested failure time models. *Encycl.Biostat.*, 7, 1998b.

James M Robins. Marginal structural models versus structural nested models as tools for causal inference. In *Statist. Mod. in Epidem., Environ., & Clin. Trials*, pages 95–133. Springer, 2000.

James M Robins and Andrea Rotnitzky. Inference for semiparametric models: Some questions and an answer-comments. *Statist. Sinica*, 11:863–885, 2001.

James M Robins, Donald Blevins, Grant Ritter, and Michael Wulfsohn. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, 3:319–336, 1992.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Am. Statist. Assoc.*, 90(429):106–121, 1995.

James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

Donald B Rubin. Bayesian inference for causal effects: The role of randomization. *Ann. Statist.*, 6:34–58, 1978.

John D Sargan. The estimation of economic relationships using instrumental variables. *Econometrica: J. Economet. Soc.*, 26:393–415, 1958.

Davey M Smith, Matthew C Strain, Simon DW Frost, Satish K Pillai, Joseph K Wong, Terri Wrin, Yang Liu, Christos J Petropolous, Eric S Daar, Susan J Little, et al. Lack of neutralizing antibody response to HIV-1 predisposes to superinfection. *Virology*, 355(1): 1–5, 2006.

Jeremy MG Taylor and Ngayee Law. Does the covariance structure matter in longitudinal modelling for the prediction of future CD4 counts? *Stat. in Med.*, 17(20):2381–2394, 1998.

Jeremy MG Taylor, WG Cumberland, and JP Sy. A stochastic model for analysis of longitudinal AIDS data. *J. Am. Statist. Assoc.*, 89(427):727–736, 1994.

Mark J Van der Laan and James M Robins. *Unified methods for censored longitudinal data and causality*. Springer Science & Business Media, 2003.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.